Chapter 2: Sampling Distribution

In a typical statistical problem, we have a random variable X of interest but its probability distribution $f(x)$ is not known. This problem can be classified in one of two ways:

1. $f(x)$ is completely unknown (Sampling Distribution).
2. The form of $f(x)$ is known but the parameter θ is unknown (Statistical Inference).

In this chapter, we will discuss the first problem and introduce some solution methods. First, let us begin with important definitions.

Random sample:

Let $X_1, X_2, \dots, X_n$ be a n independent random variables, each of which has the same probability distribution $f(x)$. Define $X_1, X_2, \dots, X_n$ to be a **random sample** of size n from the population $f(x)$ and write its joint probability distribution as

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n).$$

Statistic:

Any function of the random sample and <u>does not depend</u> upon any unknown parameter is called a **statistic**.

Sampling Distribution:

The probability distribution of a statistic is called a **sampling distribution**.

In this chapter, we studied several of the important sampling distributions of frequently used statistic. Applications of these sampling distributions to problems of statistical inference are considered throughout most of the remaining chapters.

In Chapter 1 we defined the two parameters μ and σ^2, which measure the center of location and the variability of a probability distribution, respectively. Here, we shall define some important statistics that describe corresponding measures of a random sample. The most common statistics are the sample mean and variance.

Mean and Variance:

Let $X_1, X_2, \dots, X_n$ denote a random sample of size n from a given distribution. The statistic

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

is called the **mean** of the random sample, and the statistic

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

is called the **variance** of the random sample.

Now, we should view the sampling distribution of $\bar{X}$ and S^2 as the mechanisms from which we will be able to make inference on the unknown parameters μ and σ^2.

2.1 Sampling Distribution of $\bar{X}$

Suppose that we have a population with mean μ and variance σ^2 and let $X_1, X_2, \dots, X_n$ be a random sample of size n from this population. Let the mean of the random sample be $\bar{X}$. Now, consider the following theorems of different cases of sampling distribution of $\bar{X}$.

Theorem 2.1:

Let $X_1, \dots, X_n$ be independent random variables such that, for $i = 1, \dots, n,$ X_i has a $N(\mu_i, \sigma_i^2)$ distribution. Let $Y = \sum_{i=1}^{n} a_i X_i$, where $a_1, \dots, a_n$ are constants. Then, the distribution of Y is $N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$.

Proof:

Using independent and the mgf of normal distribution, for $t \in R$, the mgf of Y is,

$$M_Y(t) = E(e^{tY}) = E\left[e^{t\sum_{i=1}^{n} a_i X_i}\right]$$

$$= \prod_{i=1}^{n} E[e^{t a_i X_i}] = \prod_{i=1}^{n} e^{a_i \mu_i t + \frac{1}{2} a_i^2 \sigma_i^2 t^2}$$

$$= e^{\sum_{i=1}^{n} a_i \mu_i t + \frac{1}{2}\sum_{i=1}^{n} a_i^2 \sigma_i^2 t^2}$$

which is the mgf of a $N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$ distribution.

Example 2.1:

Let $X_1 \sim N(3,2)$ and $X_2 \sim N(2,1)$. Find the distribution of $Y = 5X_1 - 2X_2$.

Solution:

$$E(Y) = 5E(X_1) - 2E(X_2) = 15 - 4 = 11$$

$$Var(Y) = 25E(X_1) + 4E(X_2) = 50 + 4 = 54$$

Then, the distribution of Y is obtained as

$$Y \sim N(11, 54)$$

Theorem 2.2:

If $X_1, X_2, \dots, X_n$ is a random sample from <u>any distribution</u> with mean μ and variance σ^2; then

$$\mu_{\bar{X}} = \mu \text{ and variance } \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}.$$

Proof:

Since $X_1, X_2, \dots, X_n$ is a random sample, then

$$\mu_{\bar{X}} = E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} n\,\mu = \mu.$$

$$\sigma_{\bar{x}}^2 = Var(\bar{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n\,\sigma^2 = \frac{\sigma^2}{n}.$$

Theorem 2.3:

Suppose that $X_1, X_2, \dots, X_n$ be a random sample of n observations are taken from a <u>normal population</u> with mean μ and variance σ^2. Each observation $X_i, i = 1, 2, \dots, n$, has the same normal distribution. Hence, we conclude that

1. $\bar{X}$ has a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, $\left[\text{i.e. } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\right]$.
2. $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

Proof:

Since, we know $X_1, X_2, \dots, X_n$ are independent random variables and have the same normal distribution, then they have the same mgf which is

$$M_{X_i}(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, i = 1, 2, \dots, n.$$

Now, by using the mgf transformation method (Theorem 1.8), we get

$$M_{\bar{X}}(t) = E\left(e^{\bar{X}t}\right) = E\left(e^{\frac{1}{n}\sum_{i=1}^{n} X_i\, t}\right) = E\left(e^{\frac{1}{n}(X_1+X_2+\cdots+X_n)\, t}\right)$$

$$= E\left(e^{X_1\frac{t}{n}+X_2\frac{t}{n}+\cdots+X_n\frac{t}{n}}\right) = \left(M_{X_1}\left(\frac{t}{n}\right)\right)^n, \text{ for any random variable } X_1$$

$$= \left(e^{\mu\frac{t}{n}+\frac{1}{2}\sigma^2\frac{t^2}{n^2}}\right)^n = e^{\mu t+\frac{1}{2}\frac{\sigma^2}{n}t^2}.$$

which is the mgf of the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

Theorem 2.4: Central Limit Theorem:

If $X_1, X_2, \dots, X_n$ is a random sample of size n from any distribution with mean μ and variance σ^2; if $\bar{X}$ is the mean of the random sample, then as $\boldsymbol{n} \to \infty$,

1. $\bar{X}$ has approximately a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, $\left[\text{i.e. } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\right]$.
2. $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

Example 2.2:

An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours, find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

Solution:

The sampling distribution of $\bar{X}$ will be approximately normal, with $\mu_{\bar{X}} = 800$ and $\sigma_{\bar{x}}^2 = \frac{40}{\sqrt{16}} = 10$. Then,

$$P(\bar{X} < 775) = P\left(Z < \frac{775-800}{10}\right) = P(Z < -2.5) = 0.0062.$$

Theorem 2.5:

Let $X_1, X_2, \dots, X_n$ is a random sample of size n from a normal distribution with mean μ and unknown variance σ^2, then

1. $\bar{X}$ has a t-distribution with mean μ, variance $\frac{S^2}{n}$ and $(n-1)$ degrees of freedom.
2. $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$.

Example 2.3:

A sample of 16 ten-year-old girls had a standard deviation of 12 pounds. Assume the population is normal distribution with mean weight 70 pounds. Find $P(\bar{X} > 74)$.

Solution:

We have, $\mu = 70$, $S = 12$ and $n = 16$. Then, $\bar{X}$ has a t-distribution with $n - 1 = 15$ degree of freedom. Thus,

$$P(\bar{X} > 74) = 1 - P\left(T < \frac{74-70}{12/\sqrt{16}}\right) = 1 - P(T < 1.333) = 1 - 0.9 = 0.1$$

2.2 Sampling Distributions from the Normal and Chi-Squared Distributions

In this section we introduce some sampling distributions of some important and useful random variables.

Theorem 2.6:

Let $Z \sim N(0,1)$. Then, $U = Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$ follows the chi-squared distribution with 1 degree of freedom i.e. $Z^2 \sim \chi_1^2$.

Proof:

We know that the pdf of Z is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$. Now, to find the distribution of U, use the cdf transformation method as following:

$$\begin{aligned} F_U(u) &= P(U \leq u) = P(Z^2 \leq u) = P(-\sqrt{u} \leq Z \leq \sqrt{u}) \\ &= F_Z(\sqrt{u}) - F_Z(-\sqrt{u}). \end{aligned}$$

Therefore,

$$f_U(u) = f_Z(\sqrt{u})\frac{dz}{du} - f_Z(-\sqrt{u})\frac{dz}{du}$$
$$= \frac{1}{2}u^{-\frac{1}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u} + \frac{1}{2}u^{-\frac{1}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u} = \frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)}u^{-\frac{1}{2}}e^{-\frac{u}{2}}.$$

which is the pdf of chi-squared distribution with 1 degree of freedom.

Corollary 2.1:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a normal population with mean μ and variance σ^2. If the mean of the random sample is $\bar{X}$, where $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, then

$$\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2.$$

Proof:

Left as an exercise.

Theorem 2.7:

Let $Z_1, Z_2, \dots\dots, Z_n$ be independent random variables with $Z_i = \frac{X_i-\mu_i}{\sigma_i} \sim N(0,1)$, where $X_i \sim N(\mu_i, \sigma_i)$ for each $i = 1, 2, \dots, n$. If $Y = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2$ then Y follows the chi-squared distribution with n degrees of freedom. We write $Y = \sum_{i=1}^{n} z_i^2 \sim \chi_n^2$.

Proof:

Find the moment generating function of Y. Since $Z_1, Z_2, \dots, Z_n$ are independent,

$$M_Y(t) = M_{\sum_{i=1}^{n} z_i^2}(t)$$
$$= E\left(e^{(z_1^2+z_2^2+\dots+z_n^2)t}\right)$$
$$= E\left(e^{z_1^2 t}\right).E\left(e^{z_2^2 t}\right)\dots E\left(e^{z_n^2 t}\right)$$
$$= M_{z_1^2}(t)\, M_{z_2^2}(t) \dots M_{z_n^2}(t)$$

From Theorem 2.6, each Z_i^2 follows χ_1^2 and therefore it has mgf equal to $(1-2t)^{-\frac{1}{2}}$. Conclusion:

$$M_Y(t) = \left(M_{z_1^2}(t)\right)^n = (1-2t)^{-\frac{n}{2}}, \text{ for } t > \frac{1}{2}$$

This is the mgf of chi-squared distribution with n degrees of freedom.

Corollary 2.2:

Let $X_1, X_2, \dots, X_n$ is a random sample from $N(\mu, \sigma^2)$, then $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2$.

Theorem 2.8:

If $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance of a random sample from a normal distribution with mean μ and variance σ^2, then

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Proof:

Since $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$; then we can redefine U as

$$U = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2}$$

Now, let

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}[(X_i-\mu)-(\bar{X}-\mu)]^2 \\ &= \sum_{i=1}^{n}[(X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2] \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 - 2(n\bar{X}-n\mu)(\bar{X}-\mu) + n(\bar{X}-\mu)^2 \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 - 2n(\bar{X}-\mu)^2 + n(\bar{X}-\mu)^2 \\ &= \sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2.\end{aligned}$$

Then,

$$\begin{aligned}U &= \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right] \\ &= \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 \\ &= \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2.\end{aligned}$$

Use the mgf transformation method to find the distribution of U as follows

$$M_U(t) = E(e^{U\,t}) = E\left(e^{\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2\right]t}\right)$$

Since, $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$ and $\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ are independent random variables (**Prove it**), we can get

$$M_U(t) = E\left(e^{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 t}\right) E\left(e^{-\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 t}\right)$$

$$= \frac{E\left(e^{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 t}\right)}{E\left(e^{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 t}\right)} = \frac{M_{\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}(t)}{M_{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}(t)}$$

From Corollary 2.1 and Corollary 2.2, we found that

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2 \quad \text{and} \quad \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2$$

i.e.,

$$M_{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}(t) = (1-2t)^{-\frac{n}{2}} \text{ and } M_{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}(t) = (1-2t)^{-\frac{1}{2}}$$

Then,

$$M_U(t) = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = (1-2t)^{-\frac{(n-1)}{2}}$$

which is the mgf of chi-squared distribution with $n-1$ degrees of freedom. Thus,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Theorem 2.9:

Let $X \sim \chi_n^2,\ Y \sim \chi_m^2$. If X, Y are independent then $X+Y \sim \chi_{n+m}^2$.

Proof:

Left as an exercise.

Theorem 2.10:

Let Z denote a random variable that is $Z \sim N(0,1)$; let U denote a random variable that is $U \sim \chi_k^2$ and let Z and U are independent. Then,

$$T = \frac{Z}{\sqrt{U/k}} \sim t_k$$

Proof:

Since are Z and U independent, the joint density of Z and U is given by

$$f_{Z,U}(z,u) = f_Z(z).\, f_U(u)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \frac{1}{2^{\left(\frac{k}{2}\right)}\Gamma\left(\frac{k}{2}\right)} u^{\frac{k}{2}-1} e^{-\frac{u}{2}}$$

$$= \frac{1}{2^{\left(\frac{k}{2}\right)}\Gamma\left(\frac{k}{2}\right)\sqrt{2\pi}} u^{\frac{k}{2}-1} e^{-\frac{1}{2}z^2-\frac{u}{2}}, \ u > 0, -\infty < z < \infty$$

The one-to-one transformation will be used to obtain the pdf of T. Define the random variables

$$T = \frac{Z}{\sqrt{U/k}} \text{ and } Y = U$$

Then, we can write

$$z = \frac{t\sqrt{y}}{\sqrt{k}} \text{ and } u = y$$

Therefore, the Jacobian is

$$|J| = \begin{vmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial y} \\ \frac{\partial u}{\partial t} & \frac{\partial u}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{\sqrt{y}}{\sqrt{k}} & \frac{t}{2\sqrt{ky}} \\ 0 & 1 \end{vmatrix} = \frac{\sqrt{y}}{\sqrt{k}}.$$

Thus, the joint pdf of T and Y is given by

$$f_{T,Y}(t,y) = f_{Z,U}\left(\frac{t\sqrt{y}}{\sqrt{k}}, y\right).|J|$$

$$= \frac{1}{2^{\left(\frac{k}{2}\right)}\Gamma\left(\frac{k}{2}\right)\sqrt{2\pi}} y^{\frac{k}{2}-1} e^{-\frac{yt^2}{2k}-\frac{y}{2}} \frac{\sqrt{y}}{\sqrt{k}}, \ y > 0, -\infty < t < \infty$$

The marginal pdf of T is then

$$f_T(t) = \int_0^\infty f_{T,Y}(t,y)\, dy$$

$$= \frac{1}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}} \int_0^\infty y^{\frac{k+1}{2}-1} e^{-\frac{y}{2}\left(1+\frac{t^2}{k}\right)}\, dy$$

By using gamma function, $\frac{\Gamma(\alpha)}{\beta^\alpha} = \int_0^\infty x^{\alpha-1}e^{-\beta x}\, dx$, then we get

$$f_T(t) = \frac{1}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\left(\frac{1+\frac{t^2}{k}}{2}\right)^{\frac{k+1}{2}}}$$

$$= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}; \quad -\infty < t < \infty$$

And this is the pdf of t-distribution with k degrees of freedom.

Theorem 2.11:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a $N(\mu, \sigma^2)$, where σ^2 is unknown. Then,

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Proof:

Since $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, write

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)/\left(\sigma/\sqrt{n}\right)}{\sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)\sigma^2}}}$$

From Theorem 2.3 and Theorem 2.8, we obtain

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ and } \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Then, from Theorem 2.10, we conclude that

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Theorem 2.12:

Let U and V are two independent random variables such that $U \sim \chi^2_n$ and $V \sim \chi^2_m$. Then,

$$\frac{U/n}{V/m} \sim F_{n,m}$$

where n and m are the degrees of freedom of F-distribution.

2.3 Sampling Distribution of S^2

The sample variance S^2 is given by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

From Theorem 2.8, we found that the distribution of S^2 is

$$\frac{(n-1)S^2}{\sigma^2} \sim X_{n-1}^2$$

By using this conclusion, we can calculate the mean and the variance of S^2 as follows

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \;\Rightarrow E(S^2) = \sigma^2$$

$$Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \;\Rightarrow\; Var(S^2) = \frac{2\sigma^4}{n-1}$$

Corollary 2.3:

The general derivation of the mean and the variance of the sample variance S^2 that <u>does not assume normality</u> are given by

$$E(S^2) = \sigma^2$$

$$Var(S^2) = \frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}$$

where $\mu_4 = E[(X-\mu)^4]$ is the fourth central moment of X.

2.4 Sampling Distribution of Order Statistics

In this section, the concept of order statistic will be defined and some of their properties.

Order Statistic:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a cumulative distribution function $F(x)$. Then, $Y_1 \leq Y_2 \leq \cdots \leq Y_n$, where Y_i are the X_i arranged in order of increasing degrees and are defined to be the order statistics corresponding to the random sample $X_1, X_2, \dots, X_n$.

Theorem 2.13:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a continuous cdf $F(x)$ and pdf $f(x)$; let $Y_1 \le Y_2 \le \cdots \le Y_n$ be the order statistics of this random sample. Then, the marginal pdf of any order statistic of order k, say Y_k is given by

$$f_{Y_k}(y_k) = \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1-F(y_k)]^{n-k}f(y_k), \text{ for } a < y_k < b.$$

Corollary 2.4:

As a result of Theorem 2.13, the marginal pdf of $Y_1 = min[X_1, X_2, \dots, X_n]$ and the marginal pdf of $Y_n = max[X_1, X_2, \dots, X_n]$ are, respectively, given by

$$f_{Y_1}(y_1) = n[1-F(y_1)]^{n-1}f(y_1), \text{ for } a < y_1 < b$$

$$f_{Y_n}(y_n) = n[F(y_n)]^{n-1}f(y_n), \text{ for } a < y_n < b.$$

Theorem 2.14:

Let $Y_1 \le Y_2 \le \cdots \le Y_n$ be the order statistics based on the random sample $X_1, X_2, \dots, X_n$ from a continuous distribution with pdf $f(x)$ and support (a, b). Then, the joint pdf of the order statistics is given by,

$$f(y_1, y_2, \dots, y_n) = n!\, f(y_1)f(y_2) \dots f(y_n), \quad \text{for } a < y_1 < y_2 < \cdots < y_n < b.$$

Theorem 2.15:

Let $Y_1 \le Y_2 \le \cdots \le Y_n$ be the order statistics based on the random sample $X_1, X_2, \dots, X_n$. Then, the joint pdf of any two order statistics, say $Y_r < Y_k$, is expressed in terms of cdf $F(x)$ and pdf $f(x)$ as follows

$$f_{r,k}(y_r, y_k) = \frac{n!}{(r-1)!\,(k-r-1)!\,(n-k)!}[F(y_r)]^{r-1}[F(y_k)-F(y_r)]^{k-r-1}$$

$$[1-F(y_k)]^{n-k}f(y_r)f(y_k), \; a < y_r < y_k < b$$

Example 2.4:

Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size 4 from a distribution having pdf

$$f(x) = 2x, \qquad 0 < x < 1$$

Compute:

1. $P\left(\frac{1}{2} < Y_3\right)$.
2. The joint distribution of Y_1 and Y_3.

Solution:

Here $F(x) = x^2$, provided that $0 < x < 1$, so that

1. $f_{Y_3}(y_3) = \frac{4!}{2!1!}(y_3^2)^2(1 - y_3^2)(2y_3) = 24\left(y_3^5 - y_3^7\right),\ 0 < y_3 < 1$

Thus,

$$P\left(\frac{1}{2} < Y_3\right) = \int_{\frac{1}{2}}^{1} f_{Y_3}(y_3)\, dy_3 = \int_{\frac{1}{2}}^{1} 24\left(y_3^5 - y_3^7\right) dy_3 = \frac{243}{256}.$$

2. $f_{1,3}(y_1, y_3) = \frac{4!}{0!1!\,1!}[y_1^2]^0\ [y_3^2 - y_1^2]^1\ [1 - y_3^2]^1\ 2y_1\ 2y_3$

$$= 96\ y_1\ y_3\ [y_3^2 - y_1^2]\ \ [1 - y_3^2]$$